2862–2865 Nucleic Acids Research, 2000, Vol. 28, No. 15 © 2000 Oxford University Press

Identification of sequence motifs in oligonucl otides whose presence is correlat d with antis nse activity

O. V. Matveeva*, A. D. Tsodikov[1], M. Giddings, S. M. Freier[2], J. R. Wyatt[2], A. N. Spiridonov[3], S. A. Shabalina[4], R. F. Gesteland and J. F. Atkins

Department of Human Genetics, University of Utah, 15N 2030E Room 7410, Salt Lake City, UT 84112-5330, USA, [1]Huntsman Cancer Institute, Department of Oncological Sciences, University of Utah, Salt Lake City, UT 84112, USA, [2]Isis Pharmaceuticals, Carlsbad, CA 92008, USA, [3]IHS, Ithaca, NY 14853, USA and [4]National Center for Biotechnology Information, NLM, NIH, Bethesda, MD 20814, USA

Received May 4, 2000; Revised and Accepted June 19, 2000

ABSTRACT

Design of antisense oligonucleotides targeting any mRNA can be much more efficient when several activity-enhancing motifs are included and activity-decreasing motifs are avoided. This conclusion was made after statistical analysis of data collected from >1000 experiments with phosphorothioate-modified oligonucleotides. Highly significant positive correlation between the presence of motifs CCAC, TCCC, ACTC, GCCA and CTCT in the oligonucleotide and its antisense efficiency was demonstrated. In addition, negative correlation was revealed for the motifs GGGG, ACTG, AAA and TAA. It was found that the likelihood of activity of an oligonucleotide against a desired mRNA target is sequence motif content dependent.

INTRODUCTION

Antisense oligonucleotides are gaining importance in the rapid analysis of the effects of lowered expression in tissue culture, of increasing numbers of sequenced genes, and the allure of possible therapeutic utility is being actively investigated. One important question in antisense technology is the identification of mRNA sites that can be targeted efficiently. Many experiments have shown that certain antisense oligonucleotides are more active than others in suppressing specific gene expression. A routine approach to find the most active antisense oligonucleotides involves synthesis of numerous complementary oligonucleotides (up to several dozen) for different regions of the targeted mRNA, followed by activity screening in cells (1–12). Strategies reducing the number of antisense oligonucleotides for intracellular tests should have significant benefit.

The calculated Gibbs free energy (ΔG°_{37}) values for duplex formation between an oligonucleotide and mRNA molecules correlates with oligonucleotide antisense activity (13), though hybridization affinity alone is not sufficient to ensure antisense oligonucleotide efficiency in cells (3). Systematic alignment of computer-predicted local RNA secondary structures proved to be an improvement over trial and error screening in selecting antisense oligonucleotides for inhibition of expression of the gene for intracellular adhesion molecules (14). Considerations of the predicted stabilities of antisense oligonucleotide–target-RNA duplexes and their competition with predicted secondary structures of both the targets and antisense oligonucleotides may also be valuable for antisense research (13,15).

Another theoretical strategy for identification of efficient antisense oligonucleotides arises from the finding that the motif 'TCCC' is over represented amongst the most active oligos compared to their inactive counterparts. This finding was made during analysis of published oligonucleotide sequences and in prospective experiments with TNF-α mRNA where oligos containing the TCCC motif had a much higher success rate (50%) than oligonucleotides selected by trial and error (6%) (16). The correlation between occurrence of subsequence motifs in oligonucleotides is further explored here using a database of antisense molecules from previously published experiments.

MATERIALS AND METHODS

Database

Two selection criteria were used for choosing publications from which to extract oligonucleotide sequences for inclusion in the database. First, activity of oligonucleotides must have been measured by assays that evaluated the cellular level of antisense effect on a specific mRNA or its protein product. Second, at least ten different oligonucleotides targeting the same mRNA had to be tested under identical experimental conditions. The resulting database contains the names of targeted mRNAs (genes), oligonucleotide sequences, data on their antisense activities (expressed as the ratio of levels of particular mRNA or protein measured in cells after treatment with experimental antisense versus control oligonucleotide) and literature references. The database is on the Web (http://antisense.genetics.utah.edu) and is described in more detail in a separate publication (17). Unlike previous work of this kind (18), we analyzed oligonucleotides that target different parts of mRNA rather than molecules that are complementary only to mRNA translation initiation regions. This approach permits independence from any motif bias related to initiation sequences.

*To whom correspondence should be addressed. Tel: +1 801 581 5191; Fax: +1 801 585 3910; Email: olgam@howard.genetics.utah.edu

Statistical analysis

For the database analysis, the program Oligostat (A.Tsodikov and O.Matveeva, manuscript in preparation, with the program available upon request) was created and used in combination with Excel (Microsoft, Inc.). Oligostat calculates the correlation coefficients (*t*-test) (19) of oligo activity and motif occurrence. In this work 'motif' is used for a continuous stretch of 3 or 4 nt, so that the sequence -GCCACTCT- contains six 3- and five 4-nt motifs. Correlation analysis (*t*-test) was chosen rather than the chi-square test as it avoids defining an arbitrary cut-off point in classifying oligonucleotides as active or inactive by utilizing continuous activity input. Using motifs with statistically significant correlation coefficient values ($P < 0.05$), Oligostat allows the user to create a logistic regression model (19) that relates the probability of activity with the motif content of each oligonucleotide. In addition, Oligostat performs a likelihood ratio test (19) for selection of motifs that are most significant for the model. To find motifs that are less dependent on potentially biased or inaccurate information, which might be present in any published work, 'minus one mRNA' verification was used. In this procedure, oligo subsets that target each mRNA were removed, in turn, from the database and the remaining parts of the database were analyzed to find motifs with significant correlation coefficients.

To determine the extent and reproducibility of the effect of the presence of certain motifs in increasing or decreasing the proportion of active oligonucleotides, three sets of data were used. The results obtained with each set were then compared. The first set utilized the oligonucleotides studied in the screening experiments published by ISIS Pharmaceuticals (Carlsbad, CA) (147 oligonucleotides). The second set utilized data of experiments reported by other investigators under more heterogeneous conditions (202 oligonucleotides). The third set utilized the data from unpublished experiments performed by ISIS Pharmaceuticals (908 oligonucleotides). Set 3 does not overlap with sets 1 or 2. Another test was to analyze the data with two groupings of the active oligos. One group contains oligos that decrease the level of mRNA or protein in cells to at least one-quarter of the control level and the other group combined oligos that decrease the level of mRNA or protein at least 2-fold.

RESULTS AND DISCUSSION

Correlation coefficients for motif occurrence and oligo activity were determined with the combined data sets from published experiments (sets 1 and 2). This analysis revealed several dozen motifs with significant correlation coefficient values. The list of motifs identified with Oligostat after all procedures including logistic regression, likelihood ratio test and 'minus one mRNA' test (Materials and Methods) is much shorter and includes only 10 triplet plus quadruplet motifs. Nine of these 10 motifs (all except CCGG) were also confirmed to be positive or negative predictors of oligo activity in additional testing using a database of 908 oligonucleotides (set 3 with Isis Pharmaceutical's unpublished data) (Table 1). It is noted that all motifs with positive correlation coefficients are C rich.

It is not completely clear why the presence of certain motifs correlates positively with oligonucleotide antisense activity. One consideration is that pyrimidine-rich oligonucleotides, especially those that are C-rich, are able to form the most stable DNA–RNA duplexes (20). Another is that mammalian RNase H, which is responsible for the antisense effect of phosphorothioate-modified oligonucleotides, may have some sequence preferences that could contribute to the observed bias. Preferential binding of RNase H to the A-form of heteroduplexes might be responsible for some cleavage specificity of human RNase H (21) because pyrimidine-rich sequences of antisense oligonucleotides should form an A-form duplex with RNA target sites (22). Finally, some motifs may be required for more efficient cellular uptake of oligonucleotides. The reasons for the negative correlation of some motifs are also unknown. It is likely that the motif 'GGGG' can promote self-interacting structures in oligonucleotides that would make them less available for interaction with target mRNA. Irrespective of the underlying mechanism involved, the presence of a motif identified above correlates with an increase, or decrease, of the proportion of active oligos in nearly all of each set of molecules (Fig. 1). As described in Materials and Methods, three subsets of the data were analyzed and two groupings of activity values were employed. A consistent effect was generally seen with each subdivision, though, not surprisingly because of the combination of data from different origins utilizing different assays and concentrations, some variation is also evident. One exception is the motif CCGG that was identified as the negative predictor of activity in two data subdivisions (sets 1 and 2). However, this finding was not substantiated in the third subdivision (set 3, Fig. 1B).

Table 1. List of motifs whose presence is correlated with antisense oligonucleotide activity

Motif	Correlation Coefficient	Significance
CCAC	0.3	2.6 E-09
TCCC	0.3	1.5 E-08
ACTC	0.2	4.7 E-05
GCCA	0.2	0.0015
CTCT	0.1	0.007
GGGG	-0.2	4.7 E-06
ACTG	-0.2	0.0006
TAA	-0.2	0.002
CCGG	-0.1	0.02
AAA	-0.1	0.03

Motifs in red are 'positive', their presence positively correlated with antisense oligonucleotide activity. Motifs in green are 'negative', their presence negatively correlated with activity.

It is seen from Figure 2 that the proportion of active molecules is higher for the group of oligos with several 'positive' motifs in comparison with the group of oligos with only one 'positive' motif. Combination of several 'positive' motifs in an oligonucleotide may be beneficial. The subset of oligonucleotides in the database with more than one 'positive' motif is not big enough for careful statistical analysis to address the question of whether overlapping or non-overlapping motifs are better for antisense activity.

In conclusion, the activity of antisense oligonucleotides is correlated with certain sequence motifs. Understanding the reason for this correlation will require much further work, but its existence can be used for identification of mRNA sites that are most susceptible for efficient antisense targeting.

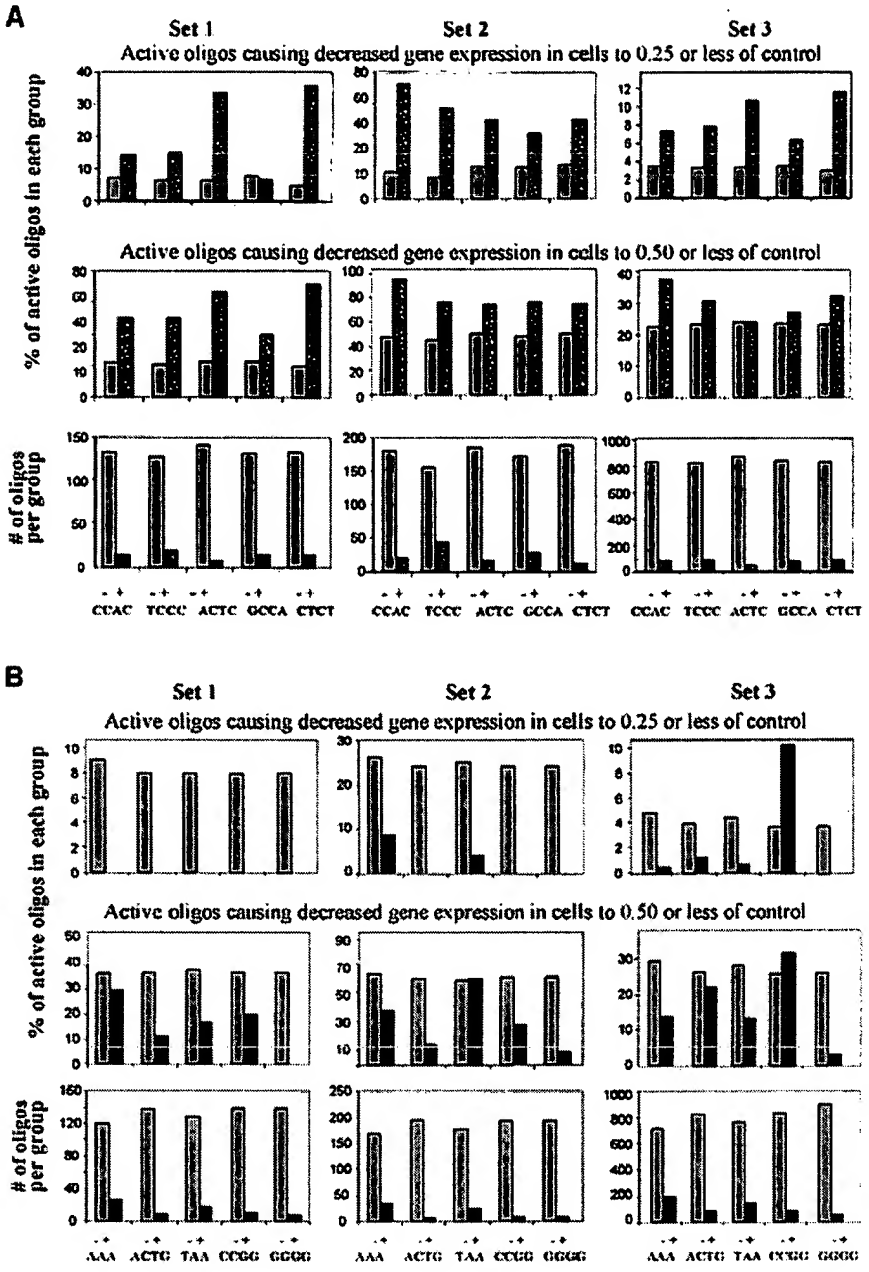

Figure 1. Positive (A) and negative (B) correlation of oligonucleotides antisense activity with the presence of some sequence motifs. Oligonucleotides from the database were categorized into groups according to the presence, or absence, of motifs that positively (A) or negatively (B) correlated with antisense activity. Red (dotted) columns represent oligonucleotides with 'positive' motifs (+). Green (hatched) columns represent oligonucleotides with 'negative' motifs (+) and blue columns represent oligonucleotides without the specified motif (–). Data in set 1 are from Isis Pharmaceutical's published work, data in set 2 are from the published work of other investigators and data in set 3 are from unpublished work of Isis Pharmaceutical.

ACKNOWLEDGEMENTS

We are grateful to Dr G. C. Tu, Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, Philadelphia, for sending us his list of selected references that was crucial for creation of our antisense oligonucleotide database. We thank Loren Miraglia from Isis Pharmaceuticals, California, for helpful discussions and comments. We also

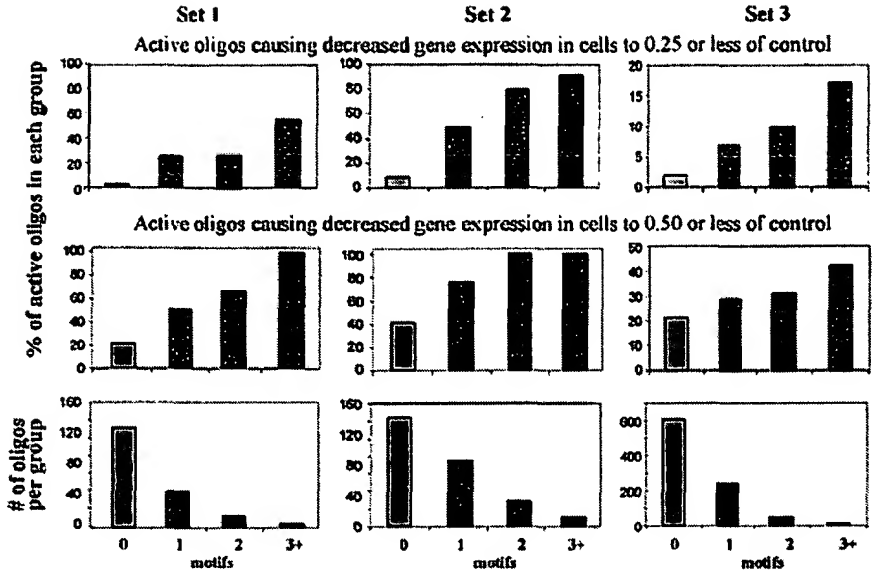

Figure 2. Correlation of oligonucleotide antisense activity with the number of positive sequence motifs. Red (spotted) columns show groups of oligos with motifs that are positively correlated with activity. Blue columns show combined groups of oligos with 'negative' or without 'positive' motifs. Data sets are described in Figure 1.

thank Dr Lex Cowsert and the Rapid Throughput Screening group at Isis Pharmaceuticals for providing the data for set 3. This work was supported in part by NIH grant (GM 48152) to J.F.A., a University of Utah Research Foundation grant (PID 9908009) to R.F.G and Cancer Center Support Grant from NIH (2P30 CA42014-12).

REFERENCES

1. Alahari,S.K., Dean,N.M., Fisher,M.H., Delong,R., Manoharan,M., Tivel,K.L. and Juliano,R.,L. (1996) *Mol. Pharmacol.*, **50**, 808–819.
2. Bennett,C.F., Condon,T.P., Grimm,S., Chan,H. and Chiang,M.Y. (1994) *J. Immunol.*, **152**, 3530–3540.
3. Chiang,M.Y., Chan,H., Zounes,M.A., Freier,S.M., Lima,W.F. and Bennett,C.F. (1991) *J. Biol. Chem.*, **266**, 18162–18171.
4. Dean,N.M., McKay,R., Condon,T.P. and Bennett,C.F. (1994) *J. Biol. Chem.*, **269**, 16416–16424.
5. Dean,N.M., McKay,R., Condon,T.P. and Bennett,C.F. (1996) *Biochem. Soc. Trans.*, **24**, 623–629.
6. Duff,J.L., Monia,B.P. and Berk,B.C. (1995) *J. Biol. Chem.*, **270**, 7161–7166.
7. Lee,C.H., Chen,H.H., Hoke,G., Jong,J.S., White,L. and Kang,Y.H. (1995) *Shock*, **4**, 1–10.
8. Lefebvre d'Hellencourt,C., Diaw,L., Cornillet,P. and Guenounou,M. (1996) *Biochim. Biophys. Acta*, **1317**, 168–174.
9. Miraglia,L., Geiger,T., Bennett,C.F. and Dean,N.M. (1996) *Int. J. Immunopharmacol.*, **18**, 227–240.
10. Monia,B.P., Johnston,J.F., Geiger,T., Muller,M. and Fabbro,D. (1996) *Nature Med.*, **2**, 668–675.
11. Stepkowski,S.M., Tu,Y., Condon,T.P. and Bennett,C.F. (1994) *J. Immunol.*, **153**, 5336–5346.
12. Stewart,A.J., Canitrot,Y., Baracchini,E., Dean,N.M., Deeley,R.G. and Cole,S.P. (1996) *Biochem. Pharmacol.*, **51**, 461–469.
13. Stull,R.A., Taylor,L.A. and Szoka,F.C.,Jr (1992) *Nucleic Acids Res.*, **20**, 3501–3508.
14. Patzel,V., Steidl,U., Kronenwett,R., Haas,R. and Sczakiel,G. (1999) *Nucleic Acids Res.*, **27**, 4328–4334.
15. Mathews,D.H., Burkard,M.E., Freier,S.M., Wyatt,J.R. and Turner,D.H. (1999) *RNA*, **5**, 1458–1469.
16. Tu,G.C., Cao,Q.N., Zhou,F. and Israel,Y. (1998) *J. Biol. Chem.*, **273**, 25125–25131.
17. Giddings,M., Matveeva,O., Atkins,J.F. and Gesteland,R.F. (2000) *Bioinformatics*, **16**, in press.
18. Smetsers,T.F., Boezeman,J.B. and Mensink,E.J. (1996) *Antisense Nucleic Acid Drug Dev.*, **6**, 63–67.
19. Sheskin,D.J. (1997) *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press, Boca Raton, FL.
20. Lesnik,E.A. and Freier,S.M. (1995) *Biochemistry*, **34**, 10807–10815.
21. Lima,W.F. and Crooke,S.T. (1997) *Biochemistry*, **3**, 390–398.
22. Ratmeyer,L., Vinayak,R., Zhong,Y.Y., Zon,G. and Wilson,W. (1994) *Biochemistry*, **33**, 5298–5304.